



University of Pennsylvania  
**ScholarlyCommons**

---

Master of Behavioral and Decision Sciences  
Capstones

Behavioral and Decision Sciences Program

---

8-9-2019

## Optimal Bias: Truth-Seeking vs. Decision-Making

Jared Peterson

Follow this and additional works at: <https://repository.upenn.edu/mbds>



Part of the [Social and Behavioral Sciences Commons](#)

---

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/mbds/17>  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## Optimal Bias: Truth-Seeking vs. Decision-Making

### Abstract

At the foundation of current statistical practices and good decision making is the idea that there is a trade-off. Due to life's uncertainties we must decide how good it is to be right and how bad it is to be wrong. Understanding the cost-benefits of being right or wrong helps us to understand systematic biases which may appear to be irrational on the surface but in actuality serve a rational function. Such an understanding may persuade decision-makers to introduce intentional bias into their decisions. Using the example of the Replication Crisis in Psychology, I will show that science is another form of decision-making which has intentional biases built in. Being consciously aware of this bias, and planning to have an optimal bias, will make for a healthier science.

### Disciplines

Social and Behavioral Sciences

---

# **Optimal Bias: Truth-Seeking vs. Decision-Making**

By Jared Peterson

## **Abstract**

At the foundation of current statistical practices and good decision making is the idea that there is a trade-off. Due to life's uncertainties we must decide how good it is to be right and how bad it is to be wrong. Understanding the cost-benefits of being right or wrong helps us to understand systematic biases which may appear to be irrational on the surface but in actuality serve a rational function. Such an understanding may persuade decision-makers to introduce intentional bias into their decisions. Using the example of the Replication Crisis in Psychology, I will show that science is another form of decision-making which has intentional biases built in. Being consciously aware of this bias, and planning to have an optimal bias, will make for a healthier science.

## The Rationality of Bias

My friend's child refuses to go downstairs at night because the tree outside the window that looks perfectly benign during the daylight becomes much more sinister after dark. One would think that millions of years of development would have trained the human perceptual system to consistently perceive a tree as merely a tree, but it is not so. This child is not alone in seeing murderers, werewolves, and aliens just outside the window.

Why does our imagination go to monsters when the lights are off? While it is true that our perceptual systems are imperfect, any explanation of the phenomenon must also explain why the reverse so rarely occurs. That is, why is it so rare to not see a face when there actually is a face? If our perceptual system was so flawed, then we would expect as many errors of one type as the other. This is not the case; with rare exceptions the errors seem to go in one direction. We mistake mannequins for people, but almost never mistake people for mannequins. Our perception is *biased*; it systematically errs in one direction.

A common way to illustrate bias is with a dart board (Kahneman, Rosenfield, Gandhi & Blaser, 2016). If the darts consistently are off in one direction (say consistently low and to the right), then there is a 'bias'. This is contrasted to 'noise', a term which means missing the bullseye in a way that lacks precision. In some situations there is bias, in others there is noise, and in others there are both. In the case of monster detection it seems to be a case of bias; we are systematically wrong in one direction. We perceive things that are not there but rarely the reverse. The question is 'why'?

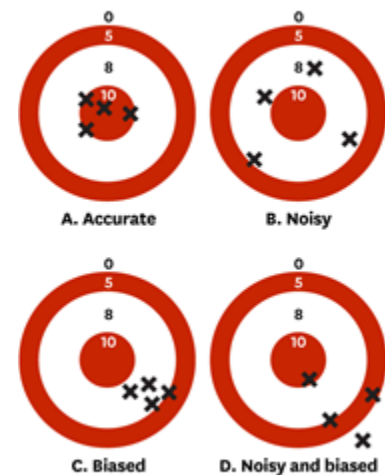
One argument that has been proposed is "Error Management Bias" (Haselton, Nettle, Murray, 2015). Though this is an evolutionary argument, the reasoning applies to many situations and is therefore a useful example.

Imagine a girl thousands of years ago settling down to sleep alongside her tribe on the open plains of Africa. She starts to drift off but then startles. Are those eyes out in the darkness? There is very little moonlight and she can't be sure. The last time she thought she saw something she woke up the whole tribe only to be embarrassed when the "eyes" were just a tree. Should she wake the tribe and risk embarrassment again?

"Better safe than sorry," she decides. She wakes the tribe, they gather their spears and walk into the darkness to find...nothing. Slightly annoyed at being awoken, the tribe settles back in to sleep.

Now imagine a second girl in the exact same situation. She sees what could be a pair of eyes in the dark. This second girl, though, is a little statistician. She has been keeping tally on a nearby rock of every time someone in the tribe thinks they see something at night. Of the last 20 times that someone in the tribe thought they saw something at night, only once did they find something out in the dark. In order to be less biased on average, meaning more correct on average, she concludes that she must have a higher level of confidence that it is indeed eyes before she wakes up the tribe. The girl goes back to sleep, proud of how rational she is for overcoming her bias. A moment later, the lions are in the camp. She may or may not survive to pass on her genes.

How Noise and Bias Affect Accuracy



SOURCE: DANIEL KAHNEMAN, ANDREW M. ROSENFELD, LINNEA GANDHI, AND TOM BLASER FROM "NOISE," OCTOBER 2016

© HBR.ORG

Two people saw something in the darkness, and both were wrong. One because she made an intentionally biased mistake (“better safe than sorry”), and the other because she wanted to be less biased on average. If the girl from our second story survives the lion attack she, too, might conclude that sometimes it is better to be safe than sorry, that is, it is better to be *biased*.

Returning to the example of dart throwing, one can imagine trying to hit an apple that is balanced on someone’s head. In this example it is very clear that a mistake of aiming too low is much worse than aiming too high. When the consequences are sufficiently severe the dart thrower may intentionally bias their shot in order to be safe, even if it means missing the dart board entirely. This bias in throwing may be explained by what psychologists call ‘Loss Aversion’. Loss Aversion is a systematic tendency to weight losses as more significant than gains (Kahneman & Tversky, 1979). If losses are significant enough then a systematic bias away from circumstances where loss can happen is the rational response. It would be better to be wrong in one direction than to make a mistake in the other direction. It would be better to mistake a tree for a lion than a lion for a tree, and better to aim too high and miss the board than too low and hit someone with the dart.

Loss Aversion is often considered irrational and not in accord with traditional economic theory but, as these examples demonstrate, the bias can be squared with traditional theories of rationality. Losses *should* loom larger than gains in many scenarios, providing nuance to the conversation around Loss Aversion (Yechiam, 2018) and showing that biases may in fact be rational in some situations.

Having established that sometimes biases, such as loss aversion, may be rational, how does one decide the optimal amount of bias? Such a question depends on the nature of the situation. When there is noise some bias may be required, but the amount of bias will be unclear until the consequences of being right or wrong have been analyzed. Once analyzed, though, it will be apparent that simple error minimization is rarely the correct strategy. Such a simple approach does not take into account the upside of being right nor the consequences of being wrong. Once these considerations are weighed one may end up biasing the results which would increase the total number of errors, but be better off for it.

### **Truth-Seeking vs. Decision-Making**

A common way to represent both error and decision making is on a 2x2 matrix. On the Y-axis there is reality and on the X-axis there is the action or conclusion of the decision-maker. This grid is used in both statistics and Game Theory. In statistics it is called a confusion-matrix (Stehman, 1997), and in Game Theory it is called the strategic-form (Bonanno, 2015). In statistics it is used to help statisticians understand how well they have identified the truth, and in game theory it is used to help decision-makers decide. To illustrate both fields and how the matrices are used we will use the example of a girl who thinks she sees a lion out in the darkness. Starting with the confusion matrix:

Confusion Matrix		
	Conclusion: It is a lion (positive)	Conclusion: It is not a lion (negative)
Reality: It is a lion	Correct ( <u>True positive</u> )	Wrong ( <u>False negative</u> )
Reality: It is not a lion	Wrong ( <u>False positive</u> )	Correct ( <u>True negative</u> )

As stated above, the X-axis represents the conclusion of the individual, and the Y-axis is possible states of reality. In situations where the individual concludes that there is a lion, this is classified as a *positive* result, and when the individual concludes there is not a lion it is a *negative* event. Additionally, if the individual is correct in their conclusion, their conclusion is *true*, or *false* if their conclusion incorrect. This means there are four possible outcomes:

**True Positives:** Correctly concluding that something is there

**True Negatives:** Correctly concluding that nothing is there

**False Positive:** Incorrectly concluding that something is there

**False Negative:** Incorrectly concluding that nothing is there

This framework is used for problems such as the lion detection example. In the absence of certainty that there is a lion there are four possible worlds, each one represented by a square in the matrix. Having established what is possible, a statistician would try to reason probabilistically about whether there is a lion or not. They could approach this problem in one of two ways; Bayesian or Frequentist.

Bayes formula is an equation for determining the probability that something is true in light of new evidence, and was developed by Reverend Bayes (1763). In the example of lion detection, an individual that used the formula would conclude that there is some probability that there is a lion; say 10%. They would come to this conclusion by thinking about the reliability of the evidence (seeing movement when there is and is not a lion) and how often lions really do approach tribes while they are asleep (called the base rate). While the formula itself might take too much time to write out when one is (possibly) facing down a lion, we can pretend for the sake of example that our ancestors were excellent mathematicians. Note that being 90% certain that there is not a lion does not inform anyone of how to handle the situation, but merely states the probability that there is a lion.

The other approach to statistics is known as frequentism. In frequentism one asks how *unlikely* evidence at least as extreme as the evidence observed would be in a situation where there was nothing (Greenland, Senn, Rothman, et al., 2016). In other words, how often does one see eyes in the shadows (or something more extreme, such as a face) when there is nothing? My own personal experience is that this happens often. Perhaps out of the last 20 times I thought I saw someone in my backyard at night only once did it turn out to be someone (my father, taking out the trash). This would be represented as follows: “ $p = .95$ ”, where .95 represents the probability of seeing a face in the shadows when there is nothing there (called a p-value).

In both the Bayesian approach and the frequentist approach we might conclude that there is not a lion since our posterior, or conclusion, with Bayes is so low (meaning low probability that there is a lion), and our p-value in frequentism so high (meaning the probability of seeing a lion when there isn't one is so high). However, while one may conclude that there is probably not a lion they may still decide to act as if there were one in order to avoid the small probability that they are wrong. The decision to act as if there were a lion despite the low probability of its presence may lead to more errors in the long run, but will help the individual avoid the one error that really matters. With this conclusion it becomes clear that what we thought was about finding the truth is actually about what decision to make even if the decision is wrong. This is where Game Theory comes in.

Game Theory is the study of decision making in a context where there is more than one decision maker. In the field, utility (total benefits minus the total costs) is assigned to outcomes that may occur which are then multiplied by the probability that those outcomes will occur (Bonanno, 2015). For example, if I make a deal that I will win \$10 if a coin lands on heads, and lose \$20 if it lands on tails, my expected utility is -\$10 because, over many iterations of the game, that would be my expected earnings. Half the time I would win \$10, and half the time I would lose \$20, for an average of -\$10. Additionally, it is important to note that utility is not always measured in dollars, but in total benefits and costs which may introduce a sense of arbitrariness. For example, I may say that I get +10 utility from being happy, and -15 for being sad. In such a situation the numbers used are arbitrary. However, to keep the analysis in the realm of rationality, the ratios between the utility values should stay consistent (e.g. 10 and -15 could be changed to 100 and -150).

Using these principles we can start to think about how to decide. We start with populating the matrix:

<b>Strategic Form Game Matrix</b>		
	<b>Wake up the tribe</b>	<b>Do not wake the tribe</b>
<b>It is a lion</b>	<i>20</i>	<i>-100</i>
<b>It is not a lion</b>	<i>-5</i>	<i>0</i>

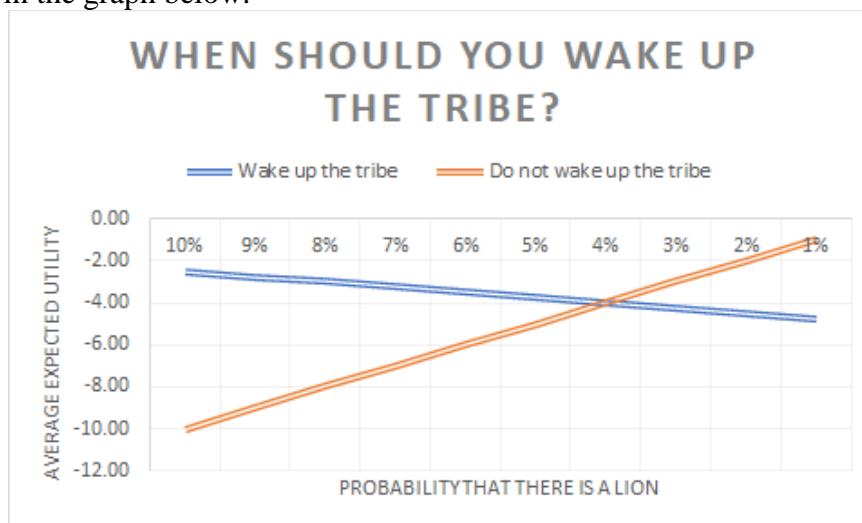
In our test case there are 4 possible outcomes; we can either wake up the tribe or not, and there may be a lion or not. I have arbitrarily decided that not waking up the tribe when there is a lion has a net negative utility of 100. From there I decided that waking up the tribe when there is not a lion is 1/20 as bad. Not waking up the tribe when there is not a lion means some extra sleep, but nothing more and so receives 0 utility. Finally, I made the decision that waking up the tribe when there a lion is four times as good as waking up the tribe when there is not a lion is bad. These estimates are arbitrary for the sake of example and should be treated so.

Next, we need an estimate of the probability that there is a lion. We will assume, as we did above, that we believe there is a 10% chance that there actually is a lion and a 90% probability that there is not a lion. One might say in such a situation that they are nearly certain that there is not a lion. Despite the certainty that there is no lion, the math comes out that one should still wake up the tribe. We discover this by multiplying the probabilities by each outcome

and then adding them up to see the average payoff. At 10% certainty the average payoff of waking the tribe is -2.5 utility, but the average payoff of not waking up the tribe is four times worse at -10. Therefore, we conclude that at 10% certainty one should still wake the tribe.

	<b>Wake up the tribe</b> Utility * probability = expected utility	<b>Do not wake the tribe</b> Utility * probability = expected utility
<b>It is a lion</b>	$20 * .10 = 2$	$-100 * .10 = -10$
<b>It is not a lion</b>	$-5 * .90 = -4.5$	$0 * .90$
<b>SUM</b>	<b>-2.5</b>	<b>-10</b>

How about if we are 95% certain there is not a lion? We should still wake the tribe. It is not until one is 96% certain that there is *not* a lion (4% certain that there is a lion) that we are indifferent as to whether we wake the tribe. At more than 96% we can go back to sleep and expect more average utility from that decision than from waking up the tribe. This intersection is shown in the graph below.



When seen through this framework it is clear why our minds tend to see things that are not there. Our ancestors needed to treat an approximate 5% probability of a lion like it was a sure thing lest they suffer rather dire consequences. This created a consistent bias in one direction where we are much more likely to see something that is not there than to miss something that is. This is not always the case across all situations, of course. There are certainly situations where the bias goes in the other direction because the utility trade-off is different, or where no bias is present. This is shown by Li, Kenrick, Griskevicius, et al., (2012) who find that Loss Aversion is domain specific, applying only to situations where such a bias would be rational. Additionally, there are even situations where we consciously make decisions about where bias should go, as in the case of criminal courts where the system is biased towards innocence. The number of situations to which we can apply this framework is quite large, and the reasoning that goes into it even larger. But what is clear is the occasionally bias should be introduced into decision making processes and that we should not make the mistake of assuming that we are truth-seeking rather



than decision-making. Whether or not there is a lion is not the most relevant question, but rather, how should one act given the evidence at hand.

### Arbitrary Numbers

In 2015, the Open Science Collaboration released a report that shook the field of psychology. They ran 100 experiments that had been published in top journals and found that only 36 of their attempted replications had the same results as the original studies that had been published. Not only that, but the studies which did replicate, meaning had the same result, had weaker results on average (Open Science Collaboration, 2015). This report and similar ones (Klein, Ratliff, Vianello, 2014; Klein, Vianello, Hasselman, et al., 2015) are part of an ongoing debate on what has been deemed “The Replication Crisis”. While the crisis is broad and applies to many fields (including, most worryingly, medicine (Begley & Ellis, 2012)), Social Psychology has borne the brunt of it in large part because of this report. A 36% replication rate was much lower than what most scientists were expecting and also lower than in other fields. It was a major blow to the field of psychology (and a few egos as well).

The Replication Crisis is a perfect practice case for thinking about bias for two reasons. One, there is an intentional bias in the statistics used in hypothesis testing, which shows that science is more like decision making than truth-seeking. And two, the actual bias goes in the opposite direction of the prescribed bias. This creates an interesting analysis of factors which might be considered when looking at bias.

In the previous section I generated rather arbitrary numbers to conclude that if someone is less than 96% certain that there is not a lion in the darkness that they should wake the tribe. These numbers were generated with very little thought. Is not detecting a lion exactly 20 times worse than waking the tribe when there is not a lion? Probably not. It is an estimation for the sake of example. However, what would be worse than such an estimation would be to use the arbitrary values chosen by someone else for a different situation, and apply it to all contexts. Unfortunately, the adoption of values set by Fisher and Cohen mean that this scenario is what has happened in psychology, as well as many other fields of science.

In 1925, Ronald Fisher set a convention that has endured for nearly 100 years (Dahiru, 2008). The convention is that for a scientist to declare that they have discovered something, that something must be *unlikely* to happen due to chance, and that this unlikelihood should be quantified. This is known as frequentism as was covered earlier. This seems to be a reasonable standard. If a coin is flipped 1000 times and it comes up heads every time, it would be reasonable to question whether the coin is, in fact, a fair coin. The unlikelihood of the event is evidence of something. Does it guarantee that the coin is not fair? No. But it very strongly implies it. Such an event should be so extremely rare that it is likely that no one in history has ever seen such an event.

What if the coin is flipped 1000 times and it comes up heads only 600 times? Or 525 times? Are those events unlikely enough that we should, at least tentatively, conclude that the coin is unfair? In the case of 1000 heads the event happens so rarely that it would probably not happen even if one tried a billion times. But what about something that happens only 20% of the time? Or 1% of the time? The cut-off to declare that something is unlikely to happen due to chance, thus implying evidence that something more than luck is at play, can be as low or high as a researcher wants. However, Fisher decided to set a standard that scientists should not declare something ‘unlikely’ unless it would happen less than 5% of the time. The number stuck. By this

standard, taken rigidly, flipping a coin and getting 526 heads would not be evidence of an unfair coin since such a result should happen 5.3% of the time. However, 527 heads only happens 4.7% of the time, and so the conclusion would be that the coin is weighted.

Fisher never meant for this metric, called the p-value as described above, to be taken so rigidly. But for many years it has been, and small random fluctuations, such as a single coin flip among 1000 such flips, have determined whether the field takes a result seriously. Many journals would not publish results unless they reached the arbitrary threshold.

A cut-off of .05 means that in 5% of the experiments where there is nothing to be found, scientists find something. Returning to our lion example, this would mean that over a period of 20 nights the tribe looks out into the darkness and thinks they see something, but nothing is ever there. Despite this, in one of those 20 nights they are convinced that the lion is there and so stay up all night guarding the camp. This cut-off at .05 has an implicit utility which states that staying up all night that one night out of twenty does not negate the times when there was a lion and it was detected plus the additional 19 nights of sleep they were able to get.

Additionally, there is another metric in science which has become a sort of default- the power of 80%. In 1988, Cohen wrote a paper and established a standard that false positives (incorrectly detecting something) are four times worse than false negatives (incorrectly finding nothing) and showed how to embed this assumption into the statistics and methodology that is used. To again return to the example of the tribe in Africa, this assumption would mean that the tribe would be indifferent in a choice between staying up all night to defend themselves against something not there, and four nights where they fail to detect a lion. This is an obviously irrational bias in the context lions.

However, this is not to disparage the bias but to praise it. Such a bias is necessary and, perhaps, not quite strong enough if one is concerned about the low replication rates in science. This is the suggestion of a paper by Benjamin et al. (2017) who argue that, instead of a p-value of .05, the threshold should instead be .005 which would require even more certainty of the presence of something before scientists declare that they have found something.

Another approach would be to not use default values at all but to instead justify the values one is using in recognition that all situations are different. This is the more rational response. As I have shown, sometimes an error in non-detecting a threat (a false negative) is many times worse than detecting something that is not there (false positive), and in other times the reverse may be true. These are questions that must be answered individually and in each separate context. This is the argument in “Justify Your Alpha” by Lakens et al. (2017). The paper’s main argument is that one should not choose any arbitrary value, but rather to justify whatever cut-off (alpha) that the authors decide to use. This cut-off could be .05 or it could be .005, but what matters is why the authors chose that value.

The authors of “Justify your Alpha” make three main arguments all of which would be relevant to anyone trying to decide on the optimal bias for a given study or set of studies. However, the third argument is especially relevant to the idea of an optimal amount of bias. The idea is this; simply requiring more confidence has downsides too. While one could require a p-value of .05 or .005 for any given study, one must also take into account the costs of getting a p-value as low as .005. As they point out, requiring that much more confidence requires 70% (or 88% depending on the type of statistical test) more subjects. Such a large increase would mean that science could only be done by well-funded organizations. Furthermore, they point out that, assuming fixed resources, fewer total studies could be done. Such a strong bias would require

more resources and is therefore not worth it according to the authors, at least not as a default rule.

Considering not only the resources for the present study, but for other potential studies is not typically how scientists would approach statistics and complicates science by quite a bit. Upon recognizing the many things which perhaps should be taken into account before one chooses an alpha, or the level of confidence required before saying something is true, it can be understandable that scientists would rather have a heuristic such as .05 than do the math and thinking required. Such laziness should be resisted as much as possible, but against should be taken into account. The time of scientists is valuable, and utility calculations not necessarily their expertise. Perhaps Cohen was right to set a p-value of .05 as a heuristic just to save the time and pain that scientists would have to go through. Or perhaps not. This paper does not set out to answer these questions, but only means to point out that such considerations should also be taken into account.

There may not be a single best answer for science. Determining the optimal bias and figuring out the ideal replication rate based on all of the relevant factors is a complicated process. It may differ by field and it may differ by study, or study type. But what is clear is that the status-quo is not good enough. Despite the strong bias towards negative results (conclusions that find nothing) the actual bias goes in the other direction. The biases introduced by Cohen and Fisher were not sufficient to outweigh all the non-statistical bias that goes into science. They can hardly be blamed for this outcome as they were trying to fight it with the values that they chose. Additionally, they never meant their values to be taken so faithfully. The problem, it would seem, is not in the statistics but in the culture and practices of scientists, and in the complications of science and statistics. This is an important take-away for anyone trying to establish an optimal bias for a given situation; there is more than probability at play and those factors must be considered as well. Before someone concludes that a coin is not fair because the statistics imply so, they should ask additional questions such as whether the coding was done correctly as well as the consequences of being wrong in either direction. In the case of science, there is the utility that the scientists themselves get from positive or negative results, which may influence how they do their analyses. Additionally, there is the pressure for scientists to find something new, novel, and interesting. These are just a few examples; there are many others.

## **Conclusion**

What is the optimal amount of bias? This question is at the center of scientific methodology and statistics. It similarly pervades other domains where there are two choices: true or false, swipe left or right, innocent or guilty, risk it or don't risk it. The problem is that one must recognize that the question is unanswerable without additional context. If there are things which might seriously maim or kill, then it is perhaps better to conclude that there is something there; "better safe than sorry". The same conclusion may be right in the context of certain experiments where there is upside and little to no downside, as may be the case when seeing how changing a few words affects responses to a survey (Bhanot, Kraft-Todd, Rand, Yoeli, 2018). But the opposite conclusion may be the case if one would invest too heavily in the uncertain results, or if an entire field is biased in the opposite direction, or if the downside is large. A bias towards non-detection is prevalent in many justice systems which are in accord with the famous Blackstone ratio; "It is better that ten guilty persons escape than that one innocent suffer."

In short, there is no single answer. One cannot conclude that false positives are always worse than false negatives, and certainly cannot conclude that they are four times worse. Additionally, one cannot conclude a false positive rate of 5% is always acceptable. Determining the optimal bias in any situation will be difficult and would require people to overcome the assumption that they are merely truth-seeking and not decision-making. In concluding that something is true, one should consider the costs of being wrong and the benefits of being right, something which is already embedded in the use of some statistical practices. Such biases should be praised for moving science in the right direction, though they should also be understood to be insufficient in light of the Replication Crisis. From this crisis one may conclude that there is too much bias in one direction, but it would be just as fair to conclude that there is insufficient bias in the opposite direction. Figuring out the optimal amount of bias requires that scientists take it case by case; the optimal bias for detecting a new cancer cure may not be the same as the optimal bias for detecting increased survey participation. Additionally, the replication rate in psychology may not need to be the same as the replication rate in physics, as each field may have a different optimal bias.

In some situations, being wrong could be fatal, and in other situations being right could be worth a lot to the decision-maker. What is important in these situations is not whether one is minimizing errors, but rather that they are maximizing utility by choosing the optimal amount of bias in the presence of uncertainty. This may lead the decision-maker to be wrong more often than not, but if done right the bias will pay-off in the long-run.

## Bibliography

- Bayes, T. (1763) An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S53Phil. Trans. R. Soc.<http://doi.org/10.1098/rstl.1763.0053>
- Begley, C. G., & Ellis, L. M. (2012, March 28). Drug development: Raise standards for preclinical cancer research. Retrieved from <https://www.nature.com/articles/483531a>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E., Berk, R., . . . Johnson, V. E. (2017, September 01). Redefine statistical significance. Retrieved from <https://www.nature.com/articles/s41562-017-0189-z>
- Bhanot, S., Kraft-Todd, G. T., Rand, D. G., & Yoeli, E. (2018, January 28). Putting Social Rewards and Identity Salience to the Test: Evidence from a Field Experiment on Teachers in Philadelphia. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3110025](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3110025)
- Bonanno, G. (2015). *Game Theory* (2nd ed.).
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Dahiru, T. (2008, June). P - value, a true test of statistical significance? A cautionary note. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111019/>
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016, May 21). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. Retrieved from <https://link.springer.com/article/10.1007/s10654-016-0149-3>
- Haselton, M. G., Nettle, D., & Murray, D. R. (2015, November 18). The Evolution of Cognitive Bias. Retrieved from <https://onlinelibrary.wiley.com/doi/full/10.1002/9781119125563.evpsych241>
- Kahneman, D., Rosenfield, A., Gandhi, L., & Blaser, T. (2018, March 06). Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making. Retrieved from <https://hbr.org/2016/10/noise>
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291. doi:10.2307/1914185
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, S., Bernstein, M. J., ... Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, 45(3), 142-152. doi: 10.1027/1864-9335/a000178
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., ... Nosek, B. A. (2018, November 19). Many Labs 2: Investigating Variation in Replicability Across Sample and Setting. <https://doi.org/10.31234/osf.io/9654g>

- Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., ... Zwaan, R. A. (2017, September 18). Justify Your Alpha. <https://doi.org/10.31234/osf.io/9s3y6>
- Li, Y. J., Kenrick, D., Griskevicius, V., & Neuberg, S. (2012). Economic decision biases and fundamental motivations: How mating and self-protection alter loss aversion. *Journal of Personality and Social Psychology*, 102(3), 550-561. <https://doi.org/10.1037/a0025844>
- Open Science Collaboration. (2015, August 28). Estimating the reproducibility of psychological science. Retrieved from <https://science.sciencemag.org/content/349/6251/aac4716>
- Stehman, S. V. (1997). *Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment*, 62(1), 77–89. doi:10.1016/s0034-4257(97)00083-7
- Yechiam, E. (2018, April 16). Acceptable losses: The debatable origins of loss aversion. Retrieved from <https://link.springer.com/article/10.1007/s00426-018-1013-8>